解答例

STEP II 2016、第 13 問

確率・統計 ・ 正規近似

STEP II 2016、問 13

先に進む前に、まず問題を読んでおこう。問題は STEP Support Programme の問題データベースにある：

step.maths.org/questions/16-s2-q13

この問題は、*Beyond STEP* の第 12.23 節「**確率から導くスターリングの公式**」で詳しく解説されている：https://beyondstep.keithmcnulty.org/ja/ch12/stirling-probability.html

鍵となる考え方

離散分布（discrete distribution）における一つの確率 $\mathrm{P}(X = k)$ は、k を中心とする幅 1 の棒の面積である。分布が正規分布（normal distribution）に近ければ、その面積はおおよそ k における正規分布の確率密度（density）に幅 1 を掛けたものになる。これを平均（mean）において適用すると、標準正規分布（standard normal distribution）の密度はそこで $\frac{1}{\sqrt{2\pi}}$ であるから、正確に分かっている確率が一つの近似式に変わる。分布をうまく選べば、その正確な確率に求めたい階乗（factorial）が含まれるようにできる。

解答

全体を通して用いる近似

W は整数値をとり、平均 μ、分散（variance）σ^2 をもち、$V \sim \mathrm{N}(\mu, \sigma^2)$ でよく近似されるとする。V の密度関数は

$$f(v) = \frac{1}{\sigma} \cdot \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2}\left(\frac{v-\mu}{\sigma}\right)^2}.$$

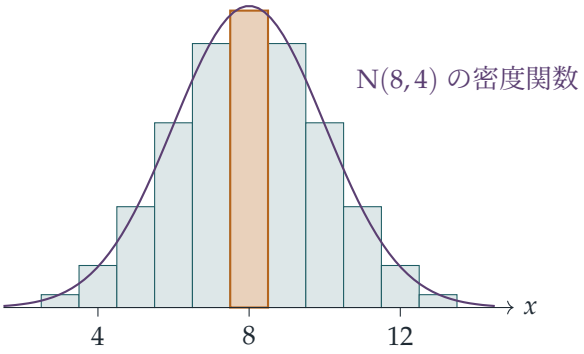
である。連続性補正（continuity correction）を行うと

$$\mathrm{P}(W = k) \approx \mathrm{P}\left(k - \tfrac{1}{2} < V < k + \tfrac{1}{2}\right) = \int_{k-1/2}^{k+1/2} f(v)\,\mathrm{d}v \approx 1 \times f(k),$$

となる。最後の近似は、区間の幅が 1 であり、σ が小さくなければ f はその区間でほとんど変化しないことによる。$k = \mu$ では指数関数の部分が $\mathrm{e}^0 = 1$ となるので

$$\mathrm{P}(W = \mu) \approx \frac{1}{\sigma\sqrt{2\pi}}. \qquad (*)$$

である。



小問 (i)

$X \sim \mathrm{B}(16, \frac{1}{2})$ に対して

$$\mathrm{E}(X) = np = 8, \qquad \mathrm{Var}(X) = np(1-p) = 16 \cdot \tfrac{1}{2} \cdot \tfrac{1}{2} = 4,$$

であるから $\sigma = 2$ である。$p = \frac{1}{2}$ なので分布は 8 に関して対称であり、n もそれなりに大きいので、$\mathrm{N}(8,4)$ はよい近似となる（図は棒グラフと正規分布の曲線を重ねて示している）。$(*)$ により、色を付けた棒の面積は

$$\mathrm{P}(X = 8) \approx \frac{1}{2\sqrt{2\pi}},$$

となり、示すべき結果が得られた。

小問 (ii)

$Y \sim \mathrm{B}(2n, \frac{1}{2})$ とする。このとき $\mathrm{E}(Y) = n$、$\mathrm{Var}(Y) = 2n \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{n}{2}$ であるから $\sigma = \sqrt{n/2}$ である。正確には

$$\mathrm{P}(Y = n) = \binom{2n}{n} \left(\tfrac{1}{2}\right)^n \left(\tfrac{1}{2}\right)^n = \frac{(2n)!}{(n!)^2\, 2^{2n}}.$$

である。n が大きければ正規近似が使え、$(*)$ より

$$\mathrm{P}(Y = n) \approx \frac{1}{\sqrt{n/2}\,\sqrt{2\pi}} = \frac{1}{\sqrt{n\pi}}.$$

を得る。二つの式を等しいとおき、$2^{2n}(n!)^2$ を掛けると

$$(2n)! \approx \frac{2^{2n}(n!)^2}{\sqrt{n\pi}},$$

となり、これが求める結果である。(小問 (i) は $n = 8$ の場合にあたる：$\sqrt{8\pi} = 2\sqrt{2\pi}$。)

小問 (iii)

$Z \sim \mathrm{Po}(n)$ とすると、$\mathrm{E}(Z) = \mathrm{Var}(Z) = n$ であり $\sigma = \sqrt{n}$ である。正確には

$$\mathrm{P}(Z = n) = \frac{\mathrm{e}^{-n} n^n}{n!}.$$

である。

> **考えてみよう** ここでポアソン分布（Poisson distribution）を正規分布で近似してよいのはなぜか。

$\mathrm{Po}(n)$ に従う確率変数は、互いに独立な n 個の $\mathrm{Po}(1)$ に従う確率変数の和と同じ分布をもつ。したがって中心極限定理（central limit theorem）により、n が大きいとき近似的に $\mathrm{N}(n, n)$ に従う。これは、λ が大きいとき $\mathrm{Po}(\lambda) \approx \mathrm{N}(\lambda, \lambda)$ という標準的な規則にほかならない。

よって $(*)$ により

$$\frac{\mathrm{e}^{-n} n^n}{n!} \approx \frac{1}{\sqrt{n}\sqrt{2\pi}} = \frac{1}{\sqrt{2\pi n}},$$

であり、変形すると

$$n! \approx \sqrt{2\pi n}\, \mathrm{e}^{-n}\, n^n,$$

となって、示すべき結果が得られた。

検算

- 小問 (i)：正確な値は $\mathrm{P}(X = 8) = \binom{16}{8}/2^{16} = 12870/65536 \approx 0.1964$ であり、一方 $\frac{1}{2\sqrt{2\pi}} \approx 0.1995$ である。誤差はおよそ 1.6% 以内である。
- 小問 (ii) で $n = 10$ のとき：$\binom{20}{10}/2^{20} \approx 0.1762$ に対して $1/\sqrt{10\pi} \approx 0.1784$ である。階乗の近似式の両辺の比は n が大きくなるにつれて 1 に近づく（$n = 100$ ではおよそ 0.9988）。
- 小問 (iii) で $n = 10$ のとき：$10! = 3\,628\,800$ であり、$\sqrt{20\pi}\,\mathrm{e}^{-10}\,10^{10} \approx 3\,598\,696$ であるから、1% 未満だけ小さい。
- 整合性：(iii) の結果を (ii) の $(2n)!$ と $n!$ の両方に代入すると、両辺とも $\sqrt{4\pi n}\,\mathrm{e}^{-2n}(2n)^{2n}$ となり、二つの近似は一致する。

よくある間違い

- 標準正規分布の密度 $\frac{1}{\sqrt{2\pi}}$ を σ で割らずに使ってしまう。$\mathrm{N}(\mu, \sigma^2)$ の平均における密度は $\frac{1}{\sigma\sqrt{2\pi}}$ である。
- 小問 (ii) で標準偏差（standard deviation）として $\sqrt{n/2}$ ではなく $\frac{n}{2}$ を使ってしまう、あるいは $\mathrm{B}(2n, \frac{1}{2})$ の分散を誤る。
- 小問 (iii) で、ポアソン分布を正規分布で近似してよい理由を述べ忘れる。

STEP のその先へ

小問 (iii) の結果はスターリングの公式（Stirling's formula）であり、この議論は中心極限定理を一点に適用した発見的な形、すなわち「局所極限定理（local limit theorem）」である。これを厳密にするには注意が必要である。分布関数（distribution function）が正規分布のものに近いと分かっても、それだけでは個々の確率が密度に近いことは言えないからである。スターリングの公式は直接証明することもでき、たとえば $\ln n! = \sum \ln k$ を $\int \ln x \, \mathrm{d}x$ と比較する方法や、ガンマ関数（gamma function）からラプラスの方法（Laplace's method）を用いる方法がある。Blitzstein と Hwang の *Introduction to Probability*（確率論入門）は中心極限定理とポアソン分布・正規分布の関係を扱っており、Grimmett と Welsh の *Probability: An Introduction*（確率論入門）はよい副読本である。

Beyond STEP 解答集：https://beyondstep.keithmcnulty.org/ja/。解答本文は CC0 1.0 である。問題そのものは © Cambridge University Press & Assessment に帰属し、ここには掲載していない。英語版から機械翻訳したものであり、まだ校閲を経ていない。